



Attorney Docket No: 014975-119

## **Substitute Specification and Abstract**



## **NUCLEIC ACID PROBES AND BROAD-RANGE PRIMERS FROM REGIONS IN TOPOISOMERASE GENES, AND METHODS IN WHICH THEY ARE USED**

### **Field of the invention**

[0001] The invention relates to nucleic acid probes and to broad-range primers that are useful in the identification of bacterial species and the diagnosis of bacterial infections. Especially, the invention relates to specific nucleic acid probes that originate from hyper-variable regions situated near the conserved sequences of topoisomerase genes of infection-causing bacteria. The invention also relates to broad-range primers originating from the conserved regions of topoisomerase genes. In addition, the invention relates to the use of these nucleic acid probes and broad-range primers in the diagnosis of bacterial infections as well as to diagnostic methods in which these nucleic acid probes and broad-range primers are used.

### **Background of the invention**

[0002] Respiratory tract infections are the most common cause for physician office visits. In Finland, approximately 13 pneumonia cases per 1000 inhabitants are diagnosed annually. Acute maxillary sinusitis and otitis are diagnosed even more frequently. For example, it is estimated that otitis accounts annually for about 200,000 cases in Finland (mainly diagnosed in children). Respiratory tract infections burden the health care system and cause significant expenses for the society. The exact costs attributable to respiratory tract infections are difficult to calculate, because the majority of these costs consist of so-called indirect costs, such as absence from work (Rautakorpi *et al.*, Scandinavian Journal of Infectious Diseases. 33(12): 920-6, 2001, Infektiaudit, ed. Eskola, Huovinen ja Valtonen 1998). Worldwide, respiratory tract infections are responsible for the deaths of millions of people, mainly children.

[0003] In addition to viruses, respiratory tract infections are caused by a variety of bacterial species : *Streptococcus pyogenes* (group A streptococcus) is an important causative agent of tonsillitis. The risk of severe complications, such as peritonsillar abscesses, is connected to untreated tonsillitis. Furthermore, sequelae of tonsillitis caused by *S. pyogenes* include rheumatic fever and glomerulonephritis that are severe, even fatal diseases. The most important causative agents of pneumonia in outpatients are *Streptococcus*

*pneumoniae* (pneumococcus), *Mycoplasma pneumoniae*, and *Chlamydia pneumoniae* from which pneumococcus is the most common and serious pathogen causing pneumonia. Less frequent pneumonia-causing bacterial species are *Legionella pneumophila* and *Coxiella burnetii*. *Mycobacterium tuberculosis*, on the other hand, causes pulmonary tuberculosis. Furthermore, pneumonia caused by rare *Mycobacterium* and *Nocardia* species is diagnosed in immunodeficient patients. In addition to pneumococcus, causative agents of maxillary sinusitis and inflammation of the middle ear (otitis media) include *Haemophilus influenzae* and *Moraxella catarrhalis*. Also *Alloioococcus otitidis* causes otitis, although less frequently.

**[0004]** At the moment the diagnosis of bacterial respiratory tract infections is mainly based on bacterial culture testing. Bacterial culture testing is, however, relatively slow, and diagnostic methods based on cultivation usually give results only days, sometimes up to weeks after sampling. Furthermore, cultivation of bacteria is not always successful under laboratory conditions. This can be a consequence of either the fact that the culture method used is not applicable for the bacterial species in question or the fact that the patient has received antimicrobial therapy before the sample is taken. In the diagnosis of pharyngitis, rapid methods based on the antigen detection are good supplementary methods to bacterial culture testing, but they work only with a limited number of pathogens (group A streptococcus). In the diagnostics of some bacterial infections (e.g., *C. pneumoniae* and *M. pneumoniae*) serological methods can also be used, but these methods give results only after several weeks from the onset of infection. Thus, they are not helpful in the acute treatment of a patient.

**[0005]** Molecular methods based on amplification and hybridization of nucleic acids attempt to solve the above-mentioned problems of bacterial culture testing. With the help of these methods the pathogen is simultaneously detected and identified, resulting in more rapid diagnostics and obviating the need for additional culture tests. Also, antibiotics do not interfere with molecular methods to such a degree as with culture testing.

**[0006]** One molecular method used in bacterial diagnostics is the so-called broad-range PCR (polymerase chain reaction) that is based on the use of broad-range primers. At the moment the most common broad-range PCR methods are based on the use of primers that recognize conserved DNA sequences of bacterial chromosomal genes encoding ribosomal RNA (16S

rDNA/rRNA or 23S rDNA/rRNA). In bacterial identification based on broad-range PCR the actual identification step is carried out by cloning and sequencing the amplified PCR product. (See e.g., EP- B1 613 502, U.S. Patent No. 6,001, 564, and U.S. Patent Application Serial No. 0020055101).

**[0007]** The broad-range bacterial PCR method has to some extent been applied to clinical bacterial diagnostics, although it is best suited for the identification of bacterial and fungal species from culture isolates, and for this purpose commercial tests, such as MicroSeq (Applied Biosystems), have also been developed. However, these tests are not widely used, because sequencing of PCR products is time-consuming and labor-intensive. Also, the assays themselves and the necessary equipment, such as sequencing instruments, are expensive, and performing the tests requires specially trained personnel. Furthermore, in polymicrobial infections extensive cloning experiments may be needed.

**[0008]** Another method used to some extent in bacterial diagnostics is multiplex PCR, which is based on the use of several specific oligonucleotides pooled in one reaction mixture. Hendolin *et al.* (Journal of Clinical Microbiology, 35: 11, 1997) used the multiplex PCR method when identifying pathogens causing otitis media.

**[0009]** Although the multiplex PCR procedure is relatively sensitive and rapid, it has some disadvantages. It is known that shorter DNA fragments are amplified more efficiently than longer DNA fragments. Therefore, if the same specimen includes two different bacterial species, the bacterial DNA of the shorter fragment is likely to be amplified more efficiently, which affects the sensitivity of the procedure. Furthermore, with multiplex PCR it is possible to identify only a few bacterial species simultaneously, because it is practically impossible to design for example over a dozen specific primer pairs so that they would be functional under the same PCR conditions. Therefore, multiplex PCR is not applicable, e.g., to the diagnostics of, e.g., respiratory tract infections in which more than ten clinically important pathogens are known.

**[0010]** Bacteria of so-called normal flora can also cause problems in microbial diagnostics based on multiplex PCR. The genome of only a few dozen bacterial species has been wholly mapped and most of these bacterial species are known disease causative agents. Thus, there is very little information about the bacteria of normal flora and their DNA sequences. This is why

the design of species-specific PCR primers and bacterial diagnostics based solely on PCR amplification are almost impossible.

**[0011]** The use of ribosomal RNA also includes some drawbacks. Distinguishing the related bacterial species from each other with the help of rRNA molecules is difficult, because the sequences of these molecules do not contain enough variable regions when they are compared with each other. And even if differences between various bacterial species may be found, these variable sites are generally divided across the whole rRNA molecule (e.g., the length of 16S rRNA is about 1500 nucleotides), which limits the use of molecular methods for diagnostics. In practice, the related bacterial species can thus be distinguished from each other only by sequencing the whole rRNA encoding gene. However, even this approach is not always sufficient to distinguish bacterial species from each other.

#### **Brief description of the invention**

**[0012]** The purpose of the present invention is to provide tools and methods, which are useful in bacterial diagnostics of infectious diseases, especially those causing respiratory tract infections, but which lack the drawbacks of bacterial diagnostics described above. In particular, the purpose of the invention is to provide novel tools and methods, which are useful in bacterial diagnostics based on molecular methods, the tools and methods being sensitive, effective, and species-specific, and being capable of identifying only the desired bacterial species. A further purpose of the invention is to provide methods, by which it is possible to diagnose infectious bacteria, especially those that cause respiratory tract infections, substantially faster than previously possible, whereby correct and effective antimicrobial therapy can be prescribed to the patient at an earlier stage of the infection so that the duration of the infection becomes shorter and the risk of potentially harmful, even life-threatening complications is reduced.

**[0013]** The present invention provides bacterial species-specific oligonucleotide probes that originate from hyper-variable regions situated near the conserved regions of genes encoding topoisomerases, especially the *gyrB/parE* genes, these hyper-variable regions differing in the base sequences significantly in various bacterial species. With these bacterial species-specific probes the genome of multiple infection-causing bacteria can be simultaneously detected and identified.

**[0014]** The invention also provides broad-range or broad-range primers that originate from the conserved regions of genes encoding topoisomerases, especially the *gyrB/parE* genes, and that efficiently amplify DNA of infection-causing bacteria even from clinical specimens, which include large amounts of foreign (non-bacterial) DNA.

**[0015]** Furthermore, the present invention provides simple, rapid, sensitive, and specific methods that overcome the drawbacks of the prior art. With these methods clinically important bacterial species can be reliably identified and diagnosed from clinical specimens or bacterial cultures.

**[0016]** The present invention relates to oligonucleotide probe sequences that hybridize under normal hybridization conditions with sequences of hyper-variable regions situated near the conserved sequences of topoisomerase genes of bacteria that cause infections, especially infections of the respiratory tract, and comprise any one of the sequences identified with SEQ ID NO: 1 to 69, and/or reverse or complementary sequences thereof, or functional fragments thereof.

**[0017]** Examples of bacterial species that cause infections, especially infections of the respiratory tract, include *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Chlamydia pneumoniae*, *Mycoplasma pneumoniae*, *Haemophilus influenzae*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Escherichia coli*, *Moraxella catarrhalis*, *Legionella pneumophila*, and *Fusobacterium necrophorum*, and examples of genes encoding topoisomerase are the gene encoding *gyrB* and *parE* proteins.

**[0018]** Preferably the length of oligonucleotide probe sequence of the invention is 15 – 30 nucleic acids, more preferably 20 to 30 and most preferably 21 to 25 nucleic acids.

**[0019]** The present invention also relates to the use of the above-mentioned oligonucleotide probe sequences in the detection, identification, or classification of bacterial species.

**[0020]** The present invention further relates to a mixture of oligonucleotide probes, which comprises any combination, and preferably all the sequences identified with SEQ ID NO: 1 to 69, and/or their reverse or complementary sequences, or functional fragments of the afore-mentioned sequences. In one preferred embodiment, the desired mixture of probes has been attached to a solid support. Preferably, an oligonucleotide probe mixture

that comprises all sequences identified with SEQ ID NO: 1 to 69, and/or their reverse or complementary sequences, has been attached onto a solid support.

**[0021]** The present invention further relates to a novel mixture of DNA primers that comprises sequences that hybridize with the sequences of conserved regions of genes encoding topoisomerases, especially with the genes encoding the *gyrB* and/or *parE* proteins of bacteria causing respiratory tract infections, and which comprise sequences identified with SEQ ID NO: 76 or 77 or their functional fragments. The present invention also relates to the use of the above mentioned mixture of primers in the amplification of topoisomerase genes, especially genes encoding the *gyrB* and *parE* proteins.

**[0022]** Furthermore, the present invention relates to a diagnostic method for detecting and identifying bacteria causing respiratory tract infections in a clinical specimen, which comprises

- a) amplifying DNA isolated from the clinical specimen using the above-mentioned mixture of primers,
- b) contacting the amplified DNA with a desired combination of the above-mentioned probes under hybridization conditions, and
- c) detecting the formation of a possible hybridization complex.

**[0023]** One preferred embodiment of the method of the invention comprises amplifying the DNA isolated from the clinical specimen by polymerase chain reaction, and contacting the amplified DNA with the bacterial species-specific oligonucleotide probes attached to the solid support.

**[0024]** In one preferred embodiment of the method of the invention, a suitably labeled nucleotide is used in the amplification of DNA isolated from the clinical specimen in order to generate a detectable target strand.

**[0025]** In another preferred embodiment of the method of the invention, the amplified and possibly labeled target DNA is contacted with a solid support, preferably with treated glass on which all the species-specific oligonucleotide probes of the invention having sequences identified with SEQ ID NO: 1 to 69 and/or their reverse or complementary sequences have been attached.

**[0026]** In a further preferred embodiment of the method of the invention, the amplified and possibly labeled target DNA is contacted with a solid support, preferably with treated glass, on which the specific oligonucleotide probes of the invention of one specified bacterium or a few specified bacteria causing respiratory tract infections have been attached, said specific oligo-

nucleotide probes having corresponding sequences identified with SEQ ID NO: 1 to 7, 8 to 14, 15 to 16, 17 to 20, 21 to 25, 26 to 29, 30 to 33, 34 to 38, 39 to 42, 43 to 47, 48 to 50, 51 to 55, 56 to 60, 61 to 65, and 66 to 69 from Table 4A and 4B and/or their reverse or complementary sequences thereof, or any suitable combination thereof.

**[0027]** The present invention further relates to a diagnostic kit for use in the diagnosis of infection-causing bacteria, especially those causing respiratory tract infections, comprising

- a) a DNA primer mixture of the invention as defined above,
- b) a mixture of bacterial species-specific oligonucleotide probe sequences, optionally attached on a solid support, of the invention as defined above,
- c) positive and optionally negative control probe sequences, and optionally
- d) reagents required in the amplification, hybridisation, purification washing, and/or detection steps.

#### **Brief description of figures**

**[0028]** Figure 1 shows examples of hyper-variable *gyrB* gene regions that are limited by the conserved sequences (*Haemophilus influenzae* SEQ ID NO:80 and *Moraxella catharralis* SEQ ID NO:81). The conserved regions have been marked with gray boxes, and they act as the annealing sites of broad-range *gyrB/parE* primers. The hyper-variable region, from which bacterial species-specific probes have been designed, is between these conserved sequences (underlined sites).

**[0029]** Figure 2 shows an example of the results of a hybridization on the surface of the microarray. The amplified DNA of the *gyrB* gene isolated from a culture isolate of *S. aureus* was used as the target strain (asymmetric Cy-5-dCTP labeled PCR product). The glass slide contained four specific *gyrB* oligonucleotide probes (Table 4A: oligonucleotide probes 21 to 24) that all bound specifically to the target strain, *S. aureus*. In addition, all 5 positive control oligonucleotides gave a detectable signal. The target strand, *S. aureus* did not bind to other oligonucleotide probes on the glass slide.

**[0030]** Figure 3 shows the PCR amplification results from Example 7, in which known primers described as specific for *Streptococcus pneumoniae* were compared in testing the bacterial species of *Streptococcus* genera and known primers described as specific for *Moraxella catarrhalis* were compared



in the testing the bacterial species of *Moraxella* genera. The results show, however, that the primers in question are not species-specific, but they also amplify the bacteria of normal flora.

#### **Detailed description of the invention**

[0031] The present invention is based on studies, which attempted to find more specific alternatives for the use of ribosomal RNA in the diagnostics of infection-causing bacteria. The study was focused on other genes that are vital for bacteria. The topoisomerases gyrase B (*gyrB* or topoisomerase II) and *parE* proteins (the other subunit of topoisomerase IV) are important molecules for bacteria, because they are needed in the packing of DNA. Certain DNA sequence fragments of these molecules have remained almost unchanged, i.e. conserved, during evolution. In this context the terms "a conserved region" or "conserved regions" refer to a region or regions of topoisomerase genes or proteins, whose nucleotide sequence or equivalently the amino acid sequence has remained nearly unchanged between different bacterial species causing respiratory tract infections. Usually these conserved regions are the most important regions for the functioning of the protein. *GyrB/ParE* molecules are, however, not as conserved as ribosomal RNA molecules. Because the molecules in question are proteins, the genes encoding these proteins include more differences at the nucleic acid level, due to the nature of their genetic code, than the genes encoding structural RNA molecules (e.g., 16S rRNA molecules). *GyrB/ParE* molecules as a whole have neither remained as unchanged as the structural RNA molecules during the evolution: short DNA fragments can be found in *gyrB/parE* molecules where the differences between various bacterial species great (including closely related bacterial species) are so great that these DNA sequences can be considered as species-specific. These sequences are hyper-variable sequences, which in this context refer to DNA sequences of the topoisomerase genes which differ in the nucleotide base sequence between different bacteria to an extent affording bacterial species-specificity and which are situated near the conserved sequences of the topoisomerase genes and optionally limited or marked out by the conserved sequences.

[0032] The above-mentioned features were utilized in the design of species-specific oligonucleotides for bacterial strains. In the design of species-specific probes (oligonucleotides) (Tables 4A and 4B) a planning strategy based on alignment was used. The *gyrB* or *parE* genes of target bacterial spe-

cies were aligned with the corresponding genes of the reference bacteria. The sequences were obtained from the EMBL public sequence database or, in the cases where such sequences were not available in public databases, they were produced by cloning the *gyrB/parE* gene sequence fragments from the bacterial species. The sequences were produced by amplifying the desired *gyrB/parE* DNA fragment from bacterial culture isolates. The desired DNA sequence fragment was then cloned and sequenced. An example of the hyper-variable *gyrB* gene region of two bacterial species, *Haemophilus influenzae* and *Moraxella catharralis*, that is bounded by conserved gene regions is presented in Figure 1. Conserved regions have been marked with gray boxes, and they act as the annealing sites of broad-range *gyrB/parE* primers. The hyper-variable region, to which species-specific probes have been designed, is between these conserved sequences (underlined sites).

**[0033]** Alignment of the sequences was performed with the BioEdit program and the ClustalW alignment algorithm. The consensus sequence of the alignments was calculated and the suitably conserved regions were identified manually. These regions refer to sequence fragments that are conserved in the genes of the target bacterial species but are not found, at least entirely, in the genes of the reference bacterial species. Oligonucleotide sequences with the suitable length (e.g., 21 – 25 nucleotides) were selected from these areas for comparison analyses. The selected oligonucleotide sequences were compared to the EMBL prokaryotic sequence database using the FastA algorithm program. The oligonucleotide sequences having at least two mismatches when compared to *gyrB/parE* sequences of non-target bacterial species were chosen for further analyses. Melting temperatures ( $T_m$ ) of oligonucleotides were calculated and the formation of hairpin structures was examined. The oligonucleotides without strong secondary structures and with  $T_m$  higher than 45 °C were selected for specificity testing. The specificity of the oligonucleotide probes was tested in laboratory conditions both with pure DNA samples isolated from various bacterial species (Example 3, Table 3) and with clinical patient samples (Example 6, Table 5).

**[0034]** The oligonucleotide probes of the present invention comprise the sequences identified with SEQ ID NO: 1 to 69, or their reverse or complementary sequences. They can be of different length, and only the desired species-specificity and functionality of these oligonucleotide probes determine their suitable length in hybridization reactions. The length of the oligonucleotide

probes is generally 15 to 30, preferably 20 to 30 and most preferably 21 to 25 nucleotides. Furthermore, the probes can be modified in different ways (e.g., modified nucleotides, such as inosine, can be included). Also, various chemical compounds or groups (e.g. amino groups) or other molecules, such as labels necessary for the detection, can be attached to the probes, or they can be entirely unmodified. The sequences of the preferred bacterial species-specific probes and their specificities are presented in Tables 4A and 4B, and they have sequences identified with SEQ ID NO: 1 to 69. Naturally, reverse and complementary sequences of these oligonucleotide sequences are equally suitable, as is obvious to a person skilled in the art. Similarly, functional fragments of the previously mentioned oligonucleotide sequences are useful as probes provided that species-specificity remains unchanged.

**[0035]** For designing the PCR primers, amino acid sequences of *gyrB* proteins of bacterial species selected from different phylogenetic groups (Table 2) were aligned with the BioEdit program using the ClustalW alignment algorithm. In alignment studies many conserved regions were found and taken as the starting point for the broad-range primer design. The *parE* protein is related to the *gyrB* protein, and the above-mentioned conserved regions can be found in both of these genes in certain bacterial species (gram-positive bacteria). The conserved amino acid sequences were reverse-translated to the corresponding nucleic acid sequences. Depending on the nature of the genetic code there were several degenerated regions in the primer sequences. On the basis of the conserved sequences several primer pairs were designed and tested in laboratory (specificity and sensitivity testing). On the basis of the laboratory tests it was concluded that the amplification of pure bacterial DNA is successful with some of the primer pairs, and that the *gyrB/parE* genes of all bacteria can be amplified with these primers. The primers in question thus act as universal or broad-range primers for the bacteria.

**[0036]** One of the primer pairs proved to have superior sensitivity as compared to the others, and clinical specimens were tested with this primer pair. However, it turned out that bacterial DNA cannot be amplified with this primer pair from clinical specimens which contain large amounts of human DNA. For this reason the primer pair was redesigned, and new primer alternatives, in terms of degeneration, were sought, which would on one hand amplify DNA from clinical specimens, and on the other hand also retain a high specificity so that *gyrB/parE* proteins of all bacteria causing respiratory tract infec-

tions could be amplified. A functional primer pair is presented in Table 1. With this primer pair, all *gyrB/parE* genes of bacteria that are phylogenetically distant from each other (Table 3) can be amplified even in the presence of large amounts of human DNA.

**Table 1. *gyrB/parE* broad-range primers**

Name of the primer	Sequence 5'→3'
gB1F	CGTCCWGGKATGTAYATHGG
gB2R	CCHACRCCRTGWAAWCCDCC

**[0037]** In the primer sequences:

W represents base A or T,

K represents base G or T,

Y represents base C or T,

H represents base A or C or T,

R represents base A or G, and

D represents base A or G or T.

**[0038]** The primer mixtures, which consist of several different primer alternatives, are thus concerned. For example, in the case of primer gB1F the mixture includes primers, in which W represents A (adenine) or T (thymine).

**[0039]** According to the invention, specific probes can be used for identification of infection-causing pathogens, especially bacteria that cause respiratory tract infections, in any suitable method, by which hybridization can be demonstrated. These methods are well known among the persons skilled in the art and they can be performed both in a solution and on a solid support that binds DNA, such as on nitrocellulose or nylon membrane, or on glass.

**[0040]** In one preferred embodiment of the method of the invention, bacterial identification is performed using the DNA microarray technology. In this context, a DNA microarray or a DNA chip refers to a small substrate on which known nucleic acid sequences have been attached in a predetermined order. If the nucleic acid fragments attached to the microarray are shorter than 100 base pairs (generally around 20 – 30 base pairs), the microarray is called an oligonucleotide array.

**[0041]** In the method of the present invention the sample to be analyzed can be a bacterial culture, a tissue fragment, a secretion sample, such as a sputum or brush sample, a blood sample, or another suitable sample, obtained from a patient suspected of suffering of an infectious disease. Especially the sample to be analyzed is a secretion sample suitable for clinical diagnostic applications.

**[0042]** DNA is isolated from the sample to be analyzed with any conventional method, such as with commercial DNA isolation kits (e.g., High Pure PCR Template Preparation Kit, Roche; NucleoSpin, BD Biosciences Clontech; or QIAamp DNA Mini-kit, Qiagen) or with traditional extraction with phenol-chloroform or equivalent organic solvents, either manually or with special devices that are suitable for performing DNA isolation. Commercial kits are preferably used because of their general availability, rapidity, and repeatability.

**[0043]** In the method of the present invention the reagents used in DNA amplification can be any reagents that are conventionally used for the amplification of DNA, and are well known among the persons skilled in the art. Suitable and cost-effective reagents, which are commercially available, include different types of Taq DNA polymerases and buffers therefor (e.g., AmpliTaqGOLD, AmpliTaqLD, DyNAzyme, TaqPlus Precision, and HotStartTaq), nucleotides or pre-prepared mixtures of nucleotides (e.g., Sigma, Applied Biosystems, Amersham Biosystems),  $MgCl_2$  (whereby a product from the manufacturer of the Taq polymerase is generally used), and Cy5-dCTP (e.g., NEN LifeSciences, Amersham Biosciences).

**[0044]** In the method of the present invention cloning can be performed with any conventionally known method, for example by using commercially available cloning kits (e.g., Qiagen PCR Cloning Kit, QIAGEN or TOPO TA Cloning Kit, Invitrogen). Sequencing of the cloning products can be performed with any sequencer suitable for this purpose (e.g., Applied Biosystems, model 373A, 377, or 3100, or Bio-Rad Sequi-Gen GT), or the products can be sequenced manually. The sequences can be analyzed manually or with sequence analysis programs designed for this purpose (e.g., Applied Biosystems Sequencer or Vector NTI Suite Version 7, InforMax).

**[0045]** The equipment used for amplification can also be any suitable device (e.g., T1 Thermocycler, Biometra, or GenAmp PCR system 2700, Applied Biosystems). Practically all devices and equipment suitable for DNA amplification can be used, and amplification can also be performed manually

by transferring reaction tubes from one temperature to another. In addition, amplification can be performed directly on a DNA microarray.

**[0046]** Purification of the PCR product can be performed with any commercial method (e.g., High Pure PCR Product Purification Kit, Roche; MicroSpin S-400, or S-300 HR Columns, Amersham Biosciences; or QIAquick PCR-purification-Kit, Qiagen) or using extraction with an organic solvent. The amplification product can also be used for the hybridization reaction as such without any further purification or extraction steps.

**[0047]** In order to form a single-stranded target strand any known digestion method can be used. These methods include, e.g., asymmetric PCR, exonuclease treatment, or the synthesis of a single-stranded target strand directly onto the microarray surface (e.g., matriXarray, Roche Applied Science). The invention also comprises applications in which a double-stranded PCR product can be used in the hybridization reaction. In the context of the present invention asymmetric PCR is the preferred method to generate a single-stranded target strand.

**[0048]** In the method of the present invention any suitable label can be used in order to produce a labeled target strand. Suitable labels include fluorescent labels (e.g., Cy5, Cy3, Cy2, TexasRed, FITC, Alexa 488, TMR, FluorX, ROX, TET, HEX), radioactive labels (e.g.,  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$ ), and chemiluminescent labels (e.g., HiLight Single-Color Kit). In the present invention the Cy5-dCTP fluorescent label (Amersham Biosciences) is preferred. The invention also comprises the applications in which no label is needed, such as those in which the detection of nucleic acids is based on electric impulse (e.g., the Motorola eSensor).

**[0049]** When hybridization takes place on a solid support, the probes used in hybridization can be attached onto the surface of the solid support by covalent or non-covalent binding. Alternatively, other chemical, electrochemical or equivalent attachment methods can be used. The substrate or support, on which the probes are attached, can be manufactured from glass, plastic, metal, nylon, nitrocellulose, polyacrylamide, silicon, or a combination of these materials, and the size of the substrate can vary from a couple of millimeters to a few centimeters. The surface of the substrate used can be treated with aminosilane or any other suitable surface treatment, such as epoxysilane, or alternatively a substrate that does not require any separate surface treatment can be used. A preferred substrate for the oligonucleotide probes is a mi-

croscopic glass slide treated with aminosilane (e.g., Genorama, Asper Biotech Ltd., Estonia).

**[0050]** The probes can be printed onto the surface of the microarray support with any commercially available arrayer that is suitable for this purpose (e.g., Qarray-mini arraying system, Lucidea Array Spotter, OmniGrid, or GeneMachines arrayer), or they can be pipetted manually onto the surface. Alternatively, the probes can be synthesized directly onto the surface by using photolithography.

**[0051]** The hybridization mixture used in hybridization can be different in its composition than what has been presented later in the working Examples, for example, the salt composition and/or the concentration can vary (e.g., 2-4xSSC or SSPE), or commercially available hybridization solutions can be used (e.g., ArrayHyb, Sigma). In addition, denaturing or stabilizing additives (e.g., formamide, DMSO, i.e. dimethyl sulfoxide) or substances that decrease non-specific binding (e.g., BSA, i.e. bovine serum albumin, or ssDNA, i.e. salmon sperm DNA) can be used in the hybridization mixture. Hybridization can be carried out in various hybridization temperatures (generally between 40-70°C), and the time needed to perform hybridization can vary, depending on the application, from a few minutes to one day. Instead of a water bath, hybridization can be carried out, e.g., in an incubator or in a special hybridization device (e.g., GeneTAC HybStation or Lucidea Slidepro Hybridizer). The post-hybridization washing steps can vary in their duration, volume, temperature, and in the composition of the washing solution, and can therefore differ from the exemplified method. The washing steps of microarray slides can also be performed with a separate device. In some cases a washing step is not necessary, because the microarray slide can be analyzed immediately after hybridization. In a preferred method a +57°C water bath is a suitable hybridization condition. The glass slides were hybridized for 12-16 hours in these conditions.

**[0052]** Microarrays or chips can be analyzed with any equipment or reader applicable for this purpose (e.g., GeneTAC UC4, GenePix Personal 4100A, or Agilent DNA Microarray Scanner). If the target strand has been marked with a fluorescent label, analysis can also be performed, e.g., with a fluorescent microscope. If a radioactive label has been used, the chip or membrane can be analyzed with autoradiography. If hybridization has been carried out on the surface of an electronic microarray and the analysis is thus based

on electronic detection, the microarray can be analyzed with special equipment designed for this purpose.

**[0053]** The genome of only a few dozen bacterial species has been entirely mapped, and most of these bacterial species are known pathogens. Thus, there is very little information about normal bacterial flora and the DNA sequences of these bacterial species. Therefore, the design of bacterial species-specific PCR primers and bacterial diagnostics based solely on PCR amplification are almost impossible.

**[0054]** The method of the present invention does not suffer from the problems of the prior art. The amplification step of the method of the present invention is very sensitive and a certain gene region (*gyrB/parE*) from phylogenetically different bacterial species were amplified efficiently with broad-range primers of the invention regardless of whether the bacterial species were gram-negative or gram-positive (Table 3). Furthermore, the PCR product is short (about 300 base pairs), which improves the effectiveness of the amplification reaction.

**[0055]** The bacterial species-specific probes have been designed for the *gyrB/parE* gene region that is considerably more variable than for example the 16S rRNA gene region, which has previously been used in bacterial diagnostics. Although bacterial species of normal flora are also amplified efficiently with broad-range primers used in the present method, no false positive reactions occur, because the probes of the invention are very bacterial species-specific and identify only those bacteria for which they have been designed. When hybridizing the target strand amplified from culture isolates of *Streptococcus pneumoniae* onto a glass slide, the target strand will hybridize only with probes specific for *Streptococcus pneumoniae* and with positive controls. On the other hand, when hybridizing the target strand amplified from bacteria of normal flora, e.g. *Streptococcus mitis*, the product will not attach to any pathogen probe; in this case only positive control probes will emit a signal (Example 7, Table 6). All specific oligonucleotide probes of the present invention were cross-tested with various bacterial species, including many bacterial species that belong to normal flora, and no cross reactions were found to take place (Figure 2, Table 6). Thus, the method of the present invention is considerably more sensitive and specific than the previously described methods of similar type.



**[0056]** The diagnostic kit of the invention can be used in the diagnosis of infection-causing bacteria, especially those causing respiratory tract infections. The kit comprises the DNA primer mixture of the invention as defined above in detail, any suitable and desirable mixture of bacterial species-specific oligonucleotide probe sequences, optionally attached on a solid support, of the invention as defined above in detail, suitable positive and optionally negative control probe sequences, and optionally reagents required in the amplification, hybridisation, purification, washing, and/or detection steps as discussed in detail above. A preferred diagnostic kit of the invention contains a DNA primer mixture comprising sequences identified with SEQ ID NO: 76 and 77, a chip containing the oligonucleotide probe sequences identified with SEQ ID NO: 1 to 69 attached thereon, positive and optionally negative controls, and optionally the reagents required to perform the method of the invention. In the following, the present invention is more precisely illustrated with the Examples. When describing the method, references have been made to different equipment, materials, temperatures, chemicals, or equivalents used in this application. These can naturally be varied in a suitable way in different applications of the invention. Therefore, the present invention and its embodiments are not limited to the Examples described below.

#### **Example 1. The design of PCR primers according to the invention**

**[0057]** For the design of PCR primers, amino acid sequences of the *gyrB* proteins (GyrB) and its related protein ParE of various bacterial species from different phylogenetic groups (Table 2) were aligned with the BioEdit program using the ClustalW alignment algorithm. Several conserved gene regions were found in the alignment of GyrB. The same conserved regions were also found from the *parE* gene of the studied gram-positive bacterial species. **[0058]** These conserved amino acid sequences were used as a starting point in the design of broad-range primers. First they were reverse-translated to the corresponding nucleic acid sequences. Because of the nature of the genetic code several degenerated sites were observed in these nucleic acid sequences. After this, based on the conserved sequences, various primer pairs were synthesized (ordered from Sigma-Genosys, England, [www.sigma-genosys.co.uk](http://www.sigma-genosys.co.uk)) and tested for specificity and sensitivity. The specificity was tested by amplifying DNA isolated from bacterial species presented in Table 4 using the method described below in Example 4. The sensitivity of the primer pair was determined by analyzing the smallest concentration of DNA isolated

from *H. Influenzae* culture isolate that could be amplified and detected with different primer pairs using the amplification and identification method described below in Example 4.

**[0059]** Based on these test results, it was noted that the amplification of pure bacterial DNA was successful for some the primer pairs, and with these primers the *gyrB/parE* genes of all studied bacterial species could be amplified. These primers functioned therefore as broad-range primers for bacteria. Sensitivity of the primers varied, and the primer pair with the highest sensitivity was used for studying the clinical specimens (otitis media samples). However, it was found out that this primer pair was not sufficiently sensitive, because it was not able to amplify bacterial DNA from clinical specimens that contained large amounts of human DNA.

**[0060]** For this reason new primer pairs that varied in terms of degeneration were synthesized (ordered from Sigma-Genosys, England, [www.sigma-genosys.co.uk](http://www.sigma-genosys.co.uk)). The specificity and the sensitivity were studied by the previously described method, both with pure bacterial DNA and with DNA isolated from clinical specimens (Table 5). A functioning primer pair was the mixture of primers, which contains the sequences

CGTCWGGKATGTAYATHGG (SEQ ID NO: 76) and  
CCHACRCCRTGWAAWCCDCC (SEQ ID NO: 77),  
which were named as gB1F (forward primer mixture) and gB2R  
(reverse primer mixture), respectively (Table 1), wherein  
W represents base A or T,  
K represents base G or T,  
Y represents base C or T,  
H represents base A or C or T,  
R represents base A or G, and  
D represents base A or G or T.

**[0061]** The conserved sequences of the first part of the *gyrB* and/or *parE* genes of all studied bacterial species were identified with this mixture of primers. It amplifies DNA from clinical samples, and has preserved sufficiently broad specificity, thus enabling the amplification of the *gyrB/parE* genes from all bacterial species causing respiratory tract infections (see Examples 6 and 7). In particular, this mixture of primers can be used to amplify the *gyrB/parE* genes of bacteria (Table 3) that are phylogenetically far from

each other even in a situation where the sample includes large amounts of human DNA.

**Table 2. Bacteria whose amino acid sequences of *gyrB* proteins were used for the alignment.**

<b>Genus</b>	<b>Species</b>
<i>Bacteroides</i>	<i>fragilis</i>
<i>Mycobacterium</i>	<i>tuberculosis</i>
<i>Streptococcus</i>	<i>pneumoniae</i>
<i>Thermotoga</i>	<i>maritima</i>
<i>Chlamydia</i>	<i>pneumoniae</i>
<i>Chlamydia</i>	<i>trachomatis</i>
<i>Borrelia</i>	<i>burgdorferi</i>
<i>Escherichia</i>	<i>coli</i>
<i>Neisseria</i>	<i>gonorrhoeae</i>
<i>Haemophilus</i>	<i>influenzae</i>
<i>Myxococcus</i>	<i>xanthus</i>
<i>Campylobacter</i>	<i>jejuni</i>
<i>Helicobacter</i>	<i>pylori</i>
<i>Bartonella</i>	<i>bacilliformis</i>
<i>Aquifex</i>	<i>aeolicus</i>

**Table 3. Bacterial strains used in sensitivity and specificity testing of primers and probes.**

<b>Bacterial species</b>	<b>Supplier's code (type of sample)</b>
<i>Hemophilus influenzae</i>	ATCC 51907D (DNA)
<i>Neisseria gonorrhoeae</i>	ATCC 53420D (DNA)
<i>Mycoplasma pneumoniae</i>	ATCC 15531D (DNA)
<i>Legionella pneumophila</i>	ATCC 33152D (DNA)
<i>Pseudomonas aeruginosa</i>	ATCC 47085D (DNA)
<i>Moraxella catarrhalis</i>	DSM 9143 (bacterial species)
<i>Staphylococcus aureus</i>	DSM 20231 (bacterial species)
<i>Streptococcus pyogenes</i>	DSM 20565 (bacterial species)
<i>Streptococcus pneumoniae</i>	DSM 20566 (bacterial species)
<i>Fusobacterium necrophorum</i>	DSM 20698 (bacterial species)
<i>Escherichia coli</i>	DSM 30083 (bacterial species)
<i>Pseudomonas aeruginosa</i>	DSM 50071 (bacterial species)
<i>Streptococcus oralis</i>	DSM 20627 (bacterial species)
<i>Streptococcus mitis</i>	DSM 12643 (bacterial species)
<i>Haemophilus parainfluenzae</i>	DSM 8978 (bacterial species)
<i>Haemophilus ducreyi</i>	DSM 8925 (bacterial species)
<i>Moraxella caviae</i>	ATCC 14659 (bacterial species)
<i>Pasteurella pneumotropica</i>	ATCC 13669 (bacterial species)
<i>Moraxella cuniculi</i>	ATCC 14688 (bacterial species)
<i>Chlamydia pneumoniae</i>	ATCC VR 1355 (bacterial species)

ATCC = American Type Culture Collection

DSM = Deutsche Sammlung von Mikroorganismen und Zellkulturen

**Example 2. Production of new sequences required in the design of oligonucleotide probes as defined by the invention by cloning**

**[0062]** *GyrB* sequences of the bacterial species *Moraxella catarrhalis*, *Fusobacterium necrophorum*, *Legionella pneumophila*, *Streptococcus mitis*, *Streptococcus oralis*, and *Haemophilus parainfluenzae* that are not available in public sequence databases were synthesized according to the general method described below in order to design oligonucleotide probes of the invention.

**[0063]** First DNA is isolated from bacterial culture isolates using the QIAamp DNA Mini kit (Qiagen, Germany). When DNA has been isolated, the desired target strand used for cloning is amplified using symmetric (conventional) polymerase chain reaction (PCR). In the first step of amplification, the reaction mixture is prepared by mixing the DNA isolated from the sample, the broad-range bacterial primers (Example 1), and the other components needed in the amplification step. Thus, the reaction mixture (25  $\mu$ l) used in cloning PCR contains 20 pmol of primer mixture gB2R, 20 pmol of primer mixture gB1F, 200  $\mu$ M of each of dATP, dGTP, dTTP, and dCTP (Sigma, USA), 1 x Hot Start Taq PCR buffer (Qiagen, Germany), which includes  $MgCl_2$  in order to achieve a final concentration of 2.8 mM, 7.5% DMSO (Amersham Pharmacia Biotech, USA), 2.5 U Hot Start Taq DNA polymerase (Qiagen, Germany), and 2.5  $\mu$ l of isolated DNA.

**[0064]** The cloning PCR is carried out in a GenAmp PCR system 2700 thermal cycler (Applied Biosystems) using the following program: a 15 min denaturation step at 95°C, 40 cycles of 1 min at 95°C, 1 min at 50°C, 1 min at 72°C, and finally a 10 min extension step at 72°C. After the PCR has been performed, the success of amplification is verified by gel electrophoresis using a 2% agarose gel containing ethidium bromide.

**[0065]** Cloning is performed immediately after PCR using a TOPO TA Cloning Kit (Invitrogen, USA). The reaction mixture for cloning contains 4  $\mu$ l of the PCR product, 1  $\mu$ l of a salt solution (1.2 M NaCl, 0.06 M  $MgCl_2$ ), and 1  $\mu$ l of TOPO vector (pCR 4-TOPO), which are mixed together in an eppendorf tube. The mixture is incubated for 5 min at room temperature, after which the solution is transferred onto ice. After this chemical transformation is performed, in which 2  $\mu$ l of cooled cloning mixture is transformed into 50  $\mu$ l of competent TOPO10 *E. coli* cells. The transformed cells are incubated for 10 min on ice. In

the next stage, a heat-shock treatment is performed. The tube containing the cells is transferred to a 42°C water bath for 30 s. After this the tube is immediately transferred onto ice and 250 µl of SOC medium at room temperature is added (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose). The tube is shaken horizontally (200 rpm) at 37°C for 1 hour. After this 20 µl of cloning mixture is spread on a pre-warmed selective LB plate (Luria-Bertani, 10% tryptone, 0.5% yeast extract, 1.0 %NaCl, 1.5% L-agar diluted in water, pH 7), which contains 50 g/ml of ampicillin. The plate is incubated overnight at 37°C. On the next day, ten colonies are chosen from the plate and sequencing PCR is performed.

**[0066]** The reaction mixture (50 µl) for sequencing PCR contains 0.4 pmol of M13 reverse (5'-CAGGAAACAGCTATGAC-3'; SEQ ID NO:78) and M13 forward (5'-GTAAACGACGGCCAG; SEQ ID NO:79) primers (provided by the kit), 150 µM of each of dATP, dGTP, dTTP, and dCTP (Sigma, USA), 1 x Hot Start Taq PCR buffer (Qiagen, Germany), 1 U Hot Start Taq DNA polymerase (Qiagen, Germany). For the amplification reaction, a small part of bacterial colony is transferred with the help of a sample stick to the PCR reaction tube.

**[0067]** The sequencing PCR is carried out in a GenAmp PCR system 2700 thermal cycler (Applied Biosystems) using the following program: a 15 min denaturation step at 95°C, 30 cycles of 1 min at 94°C, 1 min at 55°C, 1 min at 72°C, and finally a 10 min extension step at 72°C. After the PCR has been performed, the success of amplification is verified by gel electrophoresis using a 2% agarose gel that contains ethidium bromide. After this, the PCR product is purified by removing additional primers, nucleotides, buffer, and polymerase enzyme with a QIAquick PCR purification kit (Qiagen, Germany).

**[0068]** After the purification step the fragment inserted into the vector is sequenced. A 12 µl reaction mixture for the sequencing step contains 100 ng PCR product and 5 pmol of either M13 reverse or M13 forward primer. The sequencing is carried out by a BigDye Terminator Version 3.0 kit and an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, USA). The sequences are analyzed with the Vector NTI Suite Version 7 program (InforMax, USA).

**[0069]** With the above-described general method *gyrB* sequences were produced for the following bacterial species: *Moraxella catarrhalis*, *Fusobacterium necrophorum*, *Legionella pneumophila*, *Streptococcus. mitis*, *Streptococcus oralis*, and *Haemophilus parainfluenzae* (SEQ ID NO: 70 to 75).

### Example 3. Design of species-specific probes

[0070] In the design of bacterial species-specific oligonucleotides, i.e. probes, a planning strategy based on alignment was used. The *gyrB* and *parE* genes of target bacterial species were aligned with correspondent genes of a few reference bacteria (closely related bacterial species). For example, the *gyrB* gene of *Streptococcus pneumoniae* was aligned with *gyrB* genes of *Streptococcus pyogenes*, *Streptococcus mitis*, *Streptococcus oralis*, *Mycoplasma hominis*, *Staphylococcus aureus*, and *Fusobacterium necrophorum*. *S. oralis* and *S. mitis* are closely related to *S. pneumoniae*. Therefore, oligonucleotides designed for *S. pneumoniae* must not react with these normal flora bacteria. Sequences were obtained from the EMBL public sequence database or they were produced by cloning as described in Example 2.

[0071] Alignment of the sequences was performed with the BioEdit program using the ClustalW alignment algorithm. The consensus sequence of the alignments was calculated and the suitably conserved regions were identified manually. These regions refer to sequence fragments that are conserved in the genes of the target bacterial species and that are not found at least entirely from the genes of the reference bacteria. Oligonucleotide sequences with the suitable length (21 – 25 nucleotides) were selected from these areas for comparison analyses. The selected oligonucleotide sequences were compared to the EMBL prokaryotic sequence database using the FastA algorithm program. The oligonucleotide sequences having at least two mismatches when compared to *gyrB/parE* sequences of non-target bacteria were chosen for further analyses. Theoretical melting temperature ( $T_m$ ) was determined for oligonucleotides and the formation of secondary structures was studied.  $T_m$  ( $^{\circ}\text{C}$ ) was calculated with the equation

$$81.5 + 16.6 \log [\text{Na}] + 0.41(\% \text{GC}) - 0.61 (\% \text{for}) / 500 \text{N},$$

in which Na is the concentration of monovalent cations (50 M is used in calculations), %GC is the proportion of guanine and cytosine, %for is the concentration of formamide (0% is used in calculations) and N is the length of the oligonucleotide. The formation of secondary structures was studied using a program provided by Sigma-Genosys. The program can be used with the help of a web browser at the Internet address [www.sigma-genosys.co.uk/oligos/frameset.html](http://www.sigma-genosys.co.uk/oligos/frameset.html) (calculators/basic calculator). The oligonucleotides that did not form strong secondary structures and whose  $T_m$  temperature was at least  $45^{\circ}\text{C}$  were chosen for experimental specificity testing.



**[0072]** Oligonucleotide probes were synthesized and simultaneously modified from the 5' terminus (NH<sub>2</sub>-modified oligos) (Sigma-Genosys, England). The specificity of the probes was tested in the laboratory with DNA samples isolated from different bacterial species (Table 3) and from patient samples (Table 5) as described in Examples 4 and 5. Of the tested probes, those which functioned best and had the highest specificity were selected. The sequences and specificity of the bacterial species-specific probes are presented in Tables 4A and 4B.

**Table 4A. *GyrB* oligonucleotide sequences.**

Oligonucleotide/SEQ ID NO:	Specificity ( <i>gyrB</i> gene)	Sequence (5'-3')
1/1	<i>Streptococcus pneumoniae</i>	CTCAAAAGAAGGCTTCCACCATC
2/2	<i>Streptococcus pneumoniae</i>	GCCATATTCAAGTTTTTATTGAG
3/3	<i>Streptococcus pneumoniae</i>	AGCCAGATGATTGCGATTACTGTT
4/4	<i>Streptococcus pneumoniae</i>	TTGAGACCGTCTTTACAGTCCTT
5/5	<i>Streptococcus pyogenes</i>	GTTTTGCCTCTCATTTAAAGTC
6/6	<i>Streptococcus pyogenes</i>	TCTTTATTGAAGCAGATAATTCC
7/7	<i>Streptococcus pyogenes</i>	CGTTGAAACAGTTTTTACAGTCT
8/8	<i>Chlamydia pneumoniae</i>	GGTCTTCATCATCTAGTCTATGA
9/9	<i>Chlamydia pneumoniae</i>	GGTTGTAGACWACAGCATTGACG
10/10	<i>Chlamydia pneumoniae</i>	AGCCATGGCAGSTTATTGCTCTA
11/11	<i>Chlamydia pneumoniae</i>	GGATTGATGTTSGCATTTTAGAG
12/12	<i>Chlamydia pneumoniae</i>	GGGTATTGTCWTCGTAGATAATG
13/13	<i>Chlamydia pneumoniae</i>	TTATTGCTCTAGGATTGATGTT
14/14	<i>Chlamydia pneumoniae</i>	GCAATTTTAGAGSACGGGGGTATT
15/15	<i>Mycoplasma pneumoniae</i>	TCAAACCTAACCTTAAAGACAAC
16/16	<i>Mycoplasma pneumoniae</i>	TGAAACAGTGTTACGGTACTCC
17/17	<i>Haemophilus influenzae</i>	ATGATAATTCTGTATCGGTGCAA
18/18	<i>Haemophilus influenzae</i>	AGCAGAAGTTATTATGACTGTGC
19/19	<i>Haemophilus influenzae</i>	GACGATAACTCTTATAAAGTATC

20/20	<i>Haemophilus influenzae</i>	GATACCGATGACGGTACTGGTTTG
21/21	<i>Staphylococcus aureus</i>	AGTGTGGGAAATTGTCGATAATA
22/22	<i>Staphylococcus aureus</i>	TGAAGTTGTTATTGAAAAAGATAAC
23/23	<i>Staphylococcus aureus</i>	GGATTAAAGTAACGGATAACGGA
24/24	<i>Staphylococcus aureus</i>	CTGTGGAAGTTATTTTAACTGTTT
25/25	<i>Staphylococcus aureus</i>	CGACTTCAGAGAGAGGTTTGCAC
26/26	<i>Pseudomonas aeruginosa</i>	GAAATCAGCATCACCATCCATAC
27/27	<i>Pseudomonas aeruginosa</i>	ATACGGATGAGTCGACTACTGTC
28/28	<i>Pseudomonas aeruginosa</i>	GACGACAACCTACAAGGTGTC
29/29	<i>Pseudomonas aeruginosa</i>	GACACCGACGATGGCACCGGTCTG
30/30	<i>Neisseria gonorrhoeae</i>	TTCGACAACAACGCTACAAAAT
31/31	<i>Neisseria gonorrhoeae</i>	ATATGGTGTGTTGAAGTATTGGAC
32/32	<i>Neisseria gonorrhoeae</i>	GACAAAAACACGGTAACGATACA
33/33	<i>Neisseria gonorrhoeae</i>	GACAACAACGCTACAAAATCTC
34/34	<i>Escherichia coli</i>	TATTCGAGGTGGTAGATAACGCT
35/35	<i>Escherichia coli</i>	TTCACGCCGATAACTCTGTCTCT
36/36	<i>Escherichia coli</i>	CGCCGATAACTCTGTCTCTGTAC
37/37	<i>Escherichia coli</i>	GACGATAACTCCTATAAAGTGTC
38/38	<i>Escherichia coli</i>	GACACGGATGACGGCACCGGTCTG
39/39	<i>Moraxella catarrhalis</i>	AGTGCCGAGGTTATTATGACGG
40/40	<i>Moraxella catarrhalis</i>	GATGATAATTACATACAAAGTATC
41/41	<i>Moraxella catarrhalis</i>	TGTGGATATCCACCCTGAAGAAG
42/42	<i>Moraxella catarrhalis</i>	ATACCGATGATGGTACAGGCTTG
43/43	<i>Legionella pneumophila</i>	GATGATAATTCCTACAAGGTATC
44/44	<i>Legionella pneumophila</i>	AGCAGCCGAAGTCATCATGACAG
45/45	<i>Legionella pneumophila</i>	TGTAGATATTCTAAAGAAGAAG
46/46	<i>Legionella pneumophila</i>	ATACAGATGATGGAACCGGTTTG
47/47	<i>Legionella pneumophila</i>	CAGATGATGGAACCGGTTTGCAT

48/48	<i>Fusobacterium necrophorum</i>	CAACATCAGCCAGGGGATTACAT
49/49	<i>Fusobacterium necrophorum</i>	GGAATTCCAGTAGACATACATCC
50/50	<i>Fusobacterium necrophorum</i>	TGAAAATGATAACTATAAAGTGTC

**Table 4B. *ParE* oligonucleotide sequences.**

Oligonucleotide/SEQ ID NO:	Specificity ( <i>parE</i> gene)	Sequence (5'-3')
1/51	<i>Staphylococcus aureus</i>	CGGTAAACGAAATAGATGTAAACA
2/52	<i>Staphylococcus aureus</i>	AGAAGATAATGGACGTGGTATGC
3/53	<i>Staphylococcus aureus</i>	GTAAACCGACAGTCGAAGTTATC
4/54	<i>Staphylococcus aureus</i>	CAACTGATAAACGGGGATTACATC
5/55	<i>Staphylococcus aureus</i>	GGACAAGGCGGCTATAAACTTC
6/56	<i>Streptococcus pyogenes</i>	TGGAGATGATATTAAGTTGTTA
7/57	<i>Streptococcus pyogenes</i>	GTCAGTGTGGCAGATAGCGGACG
8/58	<i>Streptococcus pyogenes</i>	GGATTCCACCGTTCAAGTTATT
9/59	<i>Streptococcus pyogenes</i>	CAACAGATGCTACGGGATTGCAC
10/60	<i>Streptococcus pyogenes</i>	GGTCAGGGTGGCTACAAAACGTC
11/61	<i>Streptococcus pneumoniae</i>	TGGTGATCGTATTGATGTAACATA
12/62	<i>Streptococcus pneumoniae</i>	CTAACGGTTCAAGACCATGGACG
13/63	<i>Streptococcus pneumoniae</i>	GAATTCCAACTGTTGAGGTTATC
14/64	<i>Streptococcus pneumoniae</i>	CGACCGATGGCGCTGGTCTTCAT
15/65	<i>Streptococcus pneumoniae</i>	GGTCAAGTGGCTATAAGACATC
16/66	<i>Mycoplasma pneumoniae</i>	TGCTAATACTATTGCCGTTGTTT
17/67	<i>Mycoplasma pneumoniae</i>	ATAACTGTTAGTGACACGGTGG
18/68	<i>Mycoplasma pneumoniae</i>	AGATTTCTACGATTGACACCGTC
19/69	<i>Mycoplasma pneumoniae</i>	GATAACGATTCTTATAAGATTGC

**Example 4. Amplification of DNA isolated from patient samples**

[0073] DNAs isolated from bacterial culture isolates or clinical patient samples were amplified using the conventional method described below.

**[0074]** DNA is isolated from the sample to be analyzed (a bacterial culture or a clinical patient sample) using the QIAamp DNA Mini kit (Qiagen, Germany). When DNA has been isolated, the desired target strand is amplified using asymmetric polymerase chain reaction (PCR). In the first stage of the amplification, a reaction solution is prepared by mixing together DNA isolated from samples, broad-range gB1F and gB2R primer mixtures (Example1), and other components needed in the amplification.

**[0075]** The PCR reaction mixture contains 32 pmol of gB2R primer mixture, 8 pmol of gB1F primer mixture, 200  $\mu$ M of each of dATP, dGTP, and dTTP as well as 140  $\mu$ M dCTP (Sigma, USA), 1 x Hot Start Taq PCR buffer (Qiagen, Germany), in which  $MgCl_2$  has been added so that a final concentration is 2.8 mM, 25 nmol Cy5-AP3-dCTP (Amersham Pharmacia Biotech, USA), 7.5% DMSO (Amersham Pharmacia Biotech, USA), 2.5 U Hot Start Taq DNA polymerase (Qiagen, Germany), and 2.5  $\mu$ l isolated DNA in a total volume of 25  $\mu$ l.

**[0076]** The PCR is performed using the GenAmp PCR system 2700 thermal cycler (Applied Biosystems, USA). The following PCR program was used: a 15 min denaturation step at 95°C, 40 cycles of 1 min at 95°C, 1 min at 50°C, 1 min at 72°C, and finally a 10 min extension step at 72°C. After the PCR has been performed, the success of amplification is verified by gel electrophoresis using a 2% agarose gel that contains ethidium bromide. After this the Cy5-labeled PCR product is purified by removing additional primers, nucleotides, buffer, and polymerase enzyme with a QIAquick PCR purification kit (Qiagen, Germany).

#### **Example 5. The design and functioning of sample microarrays**

**[0077]** Oligonucleotide probes aminated at the 5' terminus (designed according to Example 3) were dissolved in 400 mM sodium carbonate buffer (pH 9.0) to a final concentration of 50  $\mu$ M. The probes were covalently attached onto aminosilane coated microscope slides (Genorama, Asper Biotech Ltd., Estonia). The transfer of the probes to the glass slides was performed with a robot developed for this purpose (OmniGrid, GeneMachines, USA) and pins (Telechem SMP3, USA). The average size of the printed probe area was 120  $\mu$ m. Moreover, positive control primers aminated at the 5' terminus were printed onto the glass slides. After printing the microarray slides were kept in ammonia vapor for 1 hour in order to attach the probes to the slides. Af-

ter the ammonia treatment they were washed three times with sterile water and dried.

**[0078]** Next, a Cy5-labeled target strand (manufactured according to Example 4) was hybridized to the microscope slide where the probes had been attached. The hybridization reaction mixture contained about 200 - 300 ng target strand, 20 x SSC (1  $\mu$ l 20xSSC contains 175.3 g NaCl and 88.2 g sodium citrate, pH is adjusted to 7.0 with HCl; the final concentration is 3.4x), 20% sodium dodecyl sulphate (SDS) (a final concentration of 0.3 %), and sterile water so that the volume of the reaction mixture was 37  $\mu$ l. First the mixture was denatured at 95°C for 3 min. After this the tubes were immediately transferred onto ice. After the mixture had cooled down, it was pipetted onto the cover slip that was placed against the glass slide on which the probes had been attached. The microarray slide was placed inside the hybridization chamber (ArrayIt, TeleChem International, USA) and the chamber was shut tight. Finally, the hybridization chamber was immersed in a water bath. The microarray slides were hybridized at 57°C for 12 - 16 hours.

**[0079]** After hybridization the microarray slides were washed in three different washing solutions in order to remove non-hybridized DNA. The washing steps were carried out as follows: in 0.1% SDS solution for 5 min at 57°C, in 0.1% SDS, 0.5xSSC washing solution for 5 min at room temperature, and in 0.06xSSC for 5 min at room temperature.

**[0080]** After the glass slides had dried, they were analyzed with a microarray scanner (Agilent DNA Microarray Scanner, Agilent, USA). If the Cy5-labeled target strand had bound to one or several probes, these spots emitted a fluorescent signal. Furthermore, positive control probes also gave a fluorescent signal.

**[0081]** An example of a hybridization result is presented in Figure 2. A *gyrB* DNA fragment (an asymmetric Cy5-labeled PCR product) isolated from a culture isolate of *Staphylococcus aureus* was used as the target strand. This example slide included four *gyrB* oligonucleotides that are specific for *Staphylococcus aureus* (Table 4B: oligonucleotides 22 - 25). All of them bound specifically the target strand of *Staphylococcus aureus* and gave a fluorescent signal. Furthermore, the five positive control oligonucleotides gave a fluorescent signal.

**Example 6. Analysis of patient samples**

**[0082]** DNA from clinical samples obtained from patients suffering from respiratory tract infections was isolated and amplified using the method described in Example 4. The samples were tested using the microarray slide (described in Example 5) on which the probes and positive control oligonucleotides (listed in Table 4A and 4B) were attached. The same samples were also analyzed with the method based on broad-range PCR, as earlier described in Nikkari *et al.* (Emerging Infectious Diseases, vol 8, nro 2, 2002, s.188 - 194) and Kotilainen *et al.* (Journal of Clinical Microbiology, vol 36, nro 8, 1998, s. 2205 - 2209).

**[0083]** Broad-range primers originating from the 16S rRNA gene area were used in the broad-range bacterial PCR method. They were named as fD1mod and 16S1RR-B primers in earlier publications. The PCR product was cloned and sequenced. The DNA sequence was compared to public sequence databases (GenBank) and the bacterial species was thus identified. The results are presented in Table 5.

**[0084]** As can be seen from Table 5, the results obtained with the method of the invention are entirely identical when compared to the conventional broad-range PCR method. The method defined by the present invention is substantially faster to perform, as results are available in approximately one day. On the other hand, it takes about one week on average to perform broad-range bacterial PCR including the time-consuming cloning and sequencing steps. Compared to culture testing, both the method of the present invention and broad-range bacterial PCR turned out to be as sensitive as cultivation and even more sensitive in some cases.

**Table 5. Comparison between the different methods using patient samples**

Clinical picture	Sample	Culture result	Broad-range PCR	The method of invention
Tonsillitis	Tissue	no information	F.N.	F.N.
Pneumonia 1)	Sputum	not cultured	no information	S.P.
Otitis media	Pus	H.I.	H.I. and M.C.	H.I. and M.C.
Otitis media	Pus	S.P.	S.P.	S.P.
Otitis media	Pus	negative	H.I.	H.I.
Otitis media	Pus	H.I.	H.I.	H.I.
Otitis media	Pus	H.I., S.P.	H.I., M.C. and S.P.	H.I., M.C. and S.P.
Maxillary sinusitis	Pus	H.I.	H.I.	H.I.
Maxillary sinusitis	Pus	negative	H.I.	H.I.

*F.N.* = *Fusobacterium necrophorum*

*S.P.* = *Streptococcus pneumoniae*

*H.I.* = *Haemophilus influenzae*

*M.C.* = *Moraxella catarrhalis*

1) Bacterial culture and broad-range PCR were not performed from the sample isolated from the patient suffering from pneumonia, only the method of the present invention was performed. With this method *Streptococcus pneumoniae* was found from the sample. This result is compatible with the clinical picture, as this bacterial species is one of the most common causative agents of pneumonia.

**Example 7. Comparison between the method of the present invention and known method that is based on specific oligonucleotides**

[0085] The identification of multiple bacterial pathogens is disturbed by bacteria of normal flora. The presence of these bacteria is necessary for the health of a human organism, and it has been estimated that hundreds of differ-

ent bacteria belong to human normal flora. Thus the bacteria of normal flora can in many cases disturb bacterial diagnostics.

**[0086]** The specificity of the method of the present invention, in which specific *gyrB/parE* oligonucleotide probes are used, was studied with respect to normal flora. Simultaneously, the method was compared to the multiplex PCR method where a mixture of broad-range and species-specific primers is used for amplification.

**[0087]** DNA was isolated from culture isolates of the normal flora bacteria *Streptococcus mitis*, *Streptococcus oralis*, *Moraxella caviae*, and *Moraxella cuniculi*. After this, isolated DNA was amplified using the method described in Example 4 and tested using the microarray slide (described in Example 5), on which specific probes (presented in Table 4A and 4B) had been attached. For comparison, specific PCR primers described in the publication of Hendolin *et al.* (Journal of Clinical Microbiology, 35; 11, 1997) were synthesized and a variant of the multiplex PCR was performed by using the method described in this publication. In this variant of the conventional multiplex PCR method the broad-range primer originating from the conserved gene region of 16S rRNA is used as one PCR primer and the mixture of primers that consists of four different species-specific primers is used as another PCR primer. Bacterial species to be diagnosed with this primer pair were *Alloicoccus otitidis*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Streptococcus pneumoniae*. Species-specific primers were designed so that the amplified PCR product differs in length depending on from which bacterial species it has been isolated. The results are presented in Table 6.

**[0088]** The results demonstrate that with the method of the present invention, in which specific oligonucleotide probes are used, only the desired bacterial species are specifically identified. Instead, specific primers defined by the known method and the multiplex PCR method cannot distinguish the desired bacteria from the studied bacteria of normal flora, and thus are not sufficiently specific (Figure 3).



**Table 6. Comparison between the method of the present invention and the multiplex PCR method**

Bacterial species	*) Probes on which they bound (specificity)	Multiplex PCR result
<i>Streptococcus pneumoniae</i>	4B: 10 - 14 ( <i>Strep.pneu</i> ) 4A: 1 - 4 ( <i>Strep.pneu</i> )	<i>Strep. pneu</i> was amplified with specific primers
<i>Streptococcus pyogenes</i>	4B: 5 - 9 ( <i>Strep.pyo</i> ) 4A: 5 - 7 ( <i>Strep. pyo</i> )	<i>Strep.pneu</i> was not amplified with specific
<i>Streptococcus oralis</i>	No binding	<del><i>Strep.pneu</i></del> <i>Strep.pneu</i> was amplified with specific primers
<i>Streptococcus mitis</i>	No binding	<i>Strep.pneu</i> was amplified with specific primers
<i>Moraxella catarrhalis</i>	4A: 39 - 42 ( <i>Morax. catarr</i> )	<i>Morax.cat.</i> was amplified with specific primers
<i>Moraxella cuniculi</i>	No binding	<i>Morax.cat.</i> was amplified with specific primers
<i>Moraxella caviae</i>	No binding	<i>Morax.cat.</i> was amplified with specific primers
<i>Neisseria gonorrhoeae</i>	4A: 30 - 33 ( <i>Neisseria gon.</i> )	<i>Morax.cat.</i> was not amplified with specific primers

\*) 4A means Table 4A and 4B Table 4B. Numbers (e.g., 10 to 14) refer to the numbers of oligonucleotides in the previously mentioned Tables. Specificity indicates for which bacterial species the oligonucleotide probes have been designed.